PACE
Policy Analysis for California Education
CORE-PACE RESEARCH PARTNERSHIP

WORKING PAPER

Can We Measure Classroom Supports for Social-Emotional Learning?

Applying Value-Added Models to Student Surveys in the CORE Districts

Teachers play a critical role in establishing classroom and school environments that contribute to students' social and emotional development. This paper explores whether we can estimate a classroom-level measure of student growth in SEL by applying value-added models to students' SEL. We analyze data from the 2016 and 2017 administrations of student self-report surveys, which contain responses from roughly 40,000 students in Grade 5 within five of California's CORE Districts. We estimate separate value-added models for each of the four SEL constructs assessed—growth mindset, self-efficacy, self-management, and social awareness—and for math and ELA academic growth. We find across-classroom-within-school variance of students' SEL outcomes, even after accounting for school-level variance. The magnitude of classroom-level impacts on students' growth in SEL appears similar to impacts on students' growth in ELA and math, although the growth models of SEL do not perform as well as growth models of academic outcomes. Results suggest that across-classroom-within-school impacts may be larger in magnitude than across-school impacts on students' SEL growth. Finally, we show that there are generally low correlations between classroom-level growth in SEL and classroom-level growth in ELA or math; however, growth mindset stands apart from the other three SEL constructs in that there is a moderately strong relationship. By assessing whether we can develop a sound approach for measuring classroom-level impacts on students' SEL, we aim to contribute to the growing body of knowledge about appropriate and innovative uses of data on students' non-cognitive and social-emotional learning.

Robert Meyer
Libby Pier
Jordan Mader
Michael Christian
Andrew Rice
Education Analytics

Susanna Loeb
Brown University

Hans Fricke
Amazon.com, Inc.

Heather Hough
Policy Analysis for California Education

VERSION: October 2019

Acknowledgements
This paper was produced as part of the CORE-PACE Research Partnership, which is focused on producing research that informs continuous improvement in the CORE districts (Fresno, Garden Grove, Long Beach, Los Angeles, Oakland, Sacramento City, San Francisco, and Santa Ana unified school districts) and policy and practice in California and beyond. We thank the CORE district administrators and leaders for their support and willingness to share their insights throughout the project. We are grateful to the Walton Family Foundation for their support of this research. Hans Fricke was the Director of Quantitative Research for PACE when producing this report and is now an economist at Amazon.com, Inc. PACE working papers are circulated for discussion and comment purposes and have not undergone the peer-review process that accompanies official PACE publications.

There is increasing recognition among educators, researchers, policymakers, and the broader public that students' social-emotional learning (SEL) is a critical component of success for both academic and life outcomes (Nagaoka, Farrington, Ehrlich, & Heath, 2015). Research indicates that educators can affect the development of these skills both directly (Durlak, Dymnicki, Taylor, Weissberg, & Schellinger, 2011) and indirectly, through the implementation of policies and practices that improve a school's culture and climate and promote positive relationships (McCormick, Cappella, O'Connor, & McClowry, 2015). Empirical evidence further suggests that classroom teachers, in particular, can play a critical role in establishing classroom and school environments that contribute to students' social and emotional development (Blazar & Kraft, 2017; Kraft, 2017; Jackson, 2014).

This increasing attention on educators' impacts on students' SEL is evident with the 2015 passage of the Every Student Succeeds Act (ESSA), which requires that states measure at least one indicator of school quality or student success that addresses measures of student engagement, educator engagement, student access to and completion of advanced coursework, postsecondary readiness, or school climate and safety. Similarly, California's Local Control Funding Formula (LCFF) and the related Local Control Accountability Plan (LCAP) process require districts to develop and report indicators of school culture and climate (California Department of Education, 2016). The state requires that school climate is measured per districts' LCAP by (i) student suspension rates (i.e., a state-wide indicator) and by (ii) locally developed measures, such as surveys of students, parents, and teachers.

The growing interest—both in California and nationwide—in using student surveys to measure student progress and school quality raises an important question about whether such surveys can reliably distinguish the impacts that different educators and schools can have on non-cognitive dimensions of success. Much in the same way that many states and districts have prioritized assessing student growth in academic achievement, rather than merely assessing student proficiency or other measures of attainment, measuring student growth in non-cognitive skills might also be more informative and equitable that relying on attainment alone. Such growth measures could then serve to highlight effective classroom supports that are successfully moving the needle to improve students' social-emotional development.

Prior research has shown that teachers have the potential to influence students' non-cognitive skills (Blazar & Kraft, 2017; Jackson, 2018). This paper aims to build upon this literature by examining whether student growth in SEL differs within schools between classrooms, and by parsing out the degree to which any classroom-level differences in students' SEL growth is attributable to a particular classroom, rather than to school-wide effects. Using survey response data from more than 40,000 fifth-grade students across five large districts in California (part of California's CORE Districts), we estimate value-added models for four SEL constructs: growth mindset, self-efficacy, self-management, and social awareness. We compare the results of these models to value-added models for math and English language arts (ELA), as well as to school-level value-added models of students' SEL. This paper serves to extend our prior work exploring school-level impacts on students' SEL (Loeb et al., 2019; Fricke et al., 2019) by further parsing out how much of the variation in students' SEL growth is due to classroom-

specific effects rather than school-general effects. In doing so, we aim to better understand the possibilities and the limitations of growth-type measures of students' SEL for different policy or practitioner applications.

Literature Review

There is no singular definition of non-cognitive or social-emotional skills, nor is there consensus on the best way to measure or assess such skills (Schweig, Baker, Hamilton, & Stecher, 2018). Indeed, AIR recently identified more than 100 different frameworks of social and emotional skills (Berg et al., 2017), and researchers have embarked on an attempt to create a taxonomy of non-cognitive skills to link terms across many of these frameworks (Jones, Bailey, Brush, Nelson, & Barnes, 2016). In addition, the RAND corporation recently identified 271 different assessments, including those used in this paper, of students' interpersonal, intrapersonal, or higher-order cognitive skills (Schweig et al., 2018). Thus, studies of students' SEL define and operationalize their outcomes of interest in many ways.

Despite the range of frameworks and assessments in the field, practitioners and researchers alike have emphasized the impact that educators have on students' social-emotional development, broadly defined. A recent survey indicated that virtually all school principals (99% of nearly 900 surveyed) believe that SEL is something that can be taught in school (DePaoli, Atwell, & Bridgeland, 2017). Other research studies have also demonstrated that this is the case (e.g., Gershenson, 2016; Jennings & DiPrete, 2010; Ladd & Sorensen, 2017; Liu & Loeb, 2018). For example, Koedel (2008) demonstrated that high school teachers significantly impact the likelihood of students' dropping out of high school, and Ladd and Sorenson (2017) showed that middle school teachers significantly reduced student absenteeism. In earlier grades, several studies point to the quality of teacher-student relationships, in particular, as uniquely predictive of students' later outcomes (Hamre & Pianta, 2001; Mashburn, et al., 2008). Thus, across students' educational trajectories, classroom teachers appear to have the potential to substantially influence students' non-cognitive development in the short term, as well as their well-being in the long-term.

Importantly, teachers who improve students' academic test performance may not be the same teachers who promote students' SEL; indeed, correlations between teachers' effects on academic assessments and on non-cognitive scores are weak (Blazar, 2018; Jackson, 2018; Liu & Loeb, 2018). Kraft (2017) found that more than one out of every four teachers in the top quartile of academic growth is in the bottom quartile of SEL growth, and Gershenson (2016) found no correlation between rankings of teachers' effects on students' absences and effects on academic achievement. Furthermore, a large proportion of teacher effects on student post-secondary outcomes, such as college enrollment, is not explained by teacher effects on student academic achievement (Chamberlain, 2013). Taken together, these studies demonstrate that there are elements of good teaching not measurable by test scores but observable in some of the non-academic skills and outcomes that we know matter for students' long-term success.

Value-Added Models of Non-Cognitive Outcomes

There is a robust body of evidence establishing the validity of using value-added models to measure the impact that teachers have on students' growth in academic achievement, as measured by student test scores (e.g., Bacher-Hicks, Chin, Kane, & Staiger, 2017; Chetty, Friedman, & Rockoff, 2014; Hanushek & Rivkin, 2010; Kane & Staiger, 2008; Kane, McCaffrey, Miller, & Staiger, 2013). From these studies, we know that teacher value-added measures vary substantially, even after taking other important indicators of teacher quality into account, such as scores on teacher licensing exams, advanced degree completion, and years of experience (Hanushek & Rivkin, 2010). Importantly, although such models have been used for purposes of teacher accountability (Koedel, Mihaly, & Rockoff, 2015), they have also been leveraged for a variety of other applications, such as evaluating the quality of professional development programs (e.g., Biancarosa, Bryk, & Dexter, 2010), assessing teacher preparation programs (e.g., Goldhaber, Liddle, & Theobald, 2013), and examining the impact of charter schools on student achievement (e.g., Betts & Tang, 2008).

However, such applications typically rely on standardized assessments of academic measures as the outcome of interest; it is only recently that value-added models have been applied to other kinds of measures. Jackson (2018) used the value-added methodology to estimate the impact of teachers on a composite of non-academic variables (i.e., student attendance, suspension rates, GPA, and on-time grade completion). His sample included more than half a million ninth- and tenth-grade students from 872 schools, taught by 5,195 English teachers and 6,854 math teachers. It is worth noting that Jackson includes both ELA and math teachers in his analyses by including in the model fixed effects of a student's "school track"—meaning that students who take the same academic courses, the same level of ELA course, and the same level of math course are all in the same school track—and assumes that the impact of other teachers averages out. This approach allows for the estimation of within-school, within-track differences, but not (i) across-classroom-within-school differences or (ii) across-school differences, as this paper aims to do; in other words, although Jackson is able to identify effects of a given academic track within a school on students' SEL, his paper does not estimate the effects of a particular *classroom* within a school on students' SEL, or the effects of different schools. Importantly, though, Jackson found that teacher effects on this index predicted students' likelihood of graduating from high school and going on to college more strongly than their effects on students' standardized test scores.

An alternative approach to relying on administrative data, such as attendance and suspension rates, is to leverage student surveys to estimate teacher value-added models of students' non-cognitive skills. Although some studies drawing a connection between teachers' pedagogy and students' social-emotional outcomes utilize self-report surveys (e.g., Blazar, 2018; Blazar & Kraft, 2017; Kraft, 2017; West, et al., 2016), leveraging such surveys to examine variation in classrooms using value-added models is a relatively novel application of such data.

Ruzek, Domina, Conley, Duncan, and Karabenick (2015) did so for 35 classrooms of roughly 2,000 seventh-grade students in seven schools, using students' self-reported

motivation orientation in math class (i.e., mastery- versus performance-oriented achievement goals); the authors found smaller variation in classroom impacts on students' achievement goals than their mathematics achievement, but found a measurable effect of teachers' abilities to mitigate a decline in students' mastery goals in math. Blazar and Kraft (2017) also estimated value-added models for 111 fourth- and fifth-grade classrooms of roughly 1,500 students using students' self-reported self-efficacy in math, happiness in class, and behaviors in class. They found large classroom effects on students' self-reported measures, comparable to what they find for classroom (i.e., teacher) effects on students' math test scores.

Taken together, value-added models at the classroom level as applied to student outcomes beyond academic assessments are relatively new, though there is some compelling evidence that they may glean useful insights into variability of classroom impacts on students' SEL. Furthermore, leveraging student self-report surveys for such models provides promising evidence that such measures may allow us to identify variability in classrooms. However, more evidence is needed to further assess the quality, strengths, and limitations of using student survey measures to estimate classroom-level impacts on students' SEL. Though there is a growing body of evidence of the reliability and validity of the self-report SEL survey measures used in this study (see Gehlbach & Hough, 2018 and Meyer, Wang, & Rice, 2018 for a comprehensive review of the evidence), this paper aims to further assess the reliability and validity of these measures in the context of classroom-level value-added models. Such evidence is crucial to accumulate as the field considers possible applications of these models, such as identifying candidate teachers for professional development opportunities focused on improving students' SEL or evaluating the quality of teacher preparation programs or teacher professional development initiatives.

Research Questions

In this paper, we seek to answer five research questions. First, we explore whether we can isolate differences in students' SEL across different classrooms within a school—in terms of their current level of SEL, their prior level of SEL, and their growth in SEL from one year to the next. These three questions allow us to first examine the feasibility and necessity of constructing growth models to estimate classroom-level impacts on students' SEL. Next, we use those models to assess the magnitude of the effects that different classrooms within a school have on students' growth in SEL. Finally, we examine whether classrooms showing high growth in SEL also show high growth in academic outcomes, in order to explore whether the classrooms promoting the highest growth in SEL are the same ones promoting the highest growth in math and ELA. Our five research questions are as follows:

1. Are there classroom-level differences in students' self-reported SEL? If so, how much of those differences can be attributed to classrooms in particular, rather than schools more generally?
2. Are there differences in students' prior levels of SEL?
3. Controlling for any differences in prior SEL, can we detect classroom-level effects of students' growth in SEL?

4. How “big” or “small” are those classroom effects?
5. Do classrooms with high SEL growth also have high growth in academic outcomes?

Data

In order to answer these questions, we rely on data from California’s CORE Districts—a consortium of eight school districts collectively serving more than one million students attending roughly 1,500 schools. In 2013, the CORE Districts secured a No Child Left Behind (NCLB) waiver that enabled them to develop a holistic school quality measurement system. One of multiple measures included in this system is school-level annual performance in four domains of students’ SEL: growth mindset, self-efficacy, self-management, and social awareness. Since the passage of ESSA (which nullified the waiver under the previous NCLB legislation), CORE Districts have continued to report school-level averages of student SEL for informational purposes, though not for direct school accountability.

CORE SEL Survey

The CORE SEL survey is administered each spring to students in grades four through 12. Students are asked to self-report their responses to items within four SEL domains: growth mindset, self-efficacy, self-management, social awareness. The surveys include between four and nine items for each of the four domains, for a total of 25 items on the survey. Each item prompts students to respond on a five-point Likert scale, indicating the extent to which the student agrees with a statement, or the extent to which a student has participated in a stated activity or experience.

West, Buckley, Krachman, and Bookman (2018) describe the four SEL domains as follows:

- *Growth mindset* is the belief that one's abilities can grow with effort. Students with a growth mindset see effort as necessary for success, embrace challenges, learn from criticism, and persist in the face of setbacks (Dweck, 2006).
- *Self-efficacy* is the belief in one's own ability to succeed in achieving an outcome or reaching a goal. Self-efficacy reflects confidence in the ability to exert control over one's motivation, behavior, and environment (Bandura, 1997).
- *Self-management* is the ability to regulate one's emotions, thoughts, and behaviors effectively in different situations. This includes managing stress, delaying gratification, motivating oneself, and setting and working toward personal and academic goals (CASEL, 2005).
- *Social awareness* is the ability to take the perspective of and empathize with others from diverse backgrounds and cultures, to understand social and ethical norms for behavior, and to recognize family, school, and community resources (CASEL, 2005).

Transforming Education (2016) and West et al. (2018) describe the process of developing the CORE SEL survey; in consultation with content experts, the survey development team curated and adapted items from multiple researcher-developed, free-to-administer SEL measures. The team conducted a pilot study with 18 schools in 2013-14, followed by a field test of the survey in 2014-15 for all CORE Districts. Starting in 2015-16, the districts began reporting school-level averages for each construct (note that they do not report growth in SEL measures).

SEL Scale Score Estimation

Measuring SEL growth using these data requires us to transform the responses to the SEL items on the student survey into a metric. We created Item Response Theory (IRT) scale scores for each of the four SEL constructs for students who responded to at least half of the survey items associated with that construct. We use a generalized partial credit model (GPCM) to produce a scale score for each of the four constructs from the responses to these items. Based on Muraki's (1992) extension of the partial credit model (Masters, 1982), GPCM can incorporate measures for which responses are on a multipoint scale in contrast to dichotomous items. Meyer, Wang, & Rice (2018); West et al. (2018); and Gehlbach & Hough (2018) describe the properties of the SEL scale score measures in more detail.

Analysis Sample

In this paper, we draw upon self-report SEL survey data collected from approximately 44,000 students within 3,622 classrooms in 724 schools within one of five participating CORE Districts. We focus on students in fifth grade, since these students are frequently in self-contained classrooms, with one teacher of record for a single group of students. By first establishing whether we can estimate classroom-level effects of students' SEL for self-contained classrooms, we can build a statistical foundation upon which we could later explore generalizing to more complex teacher-student links (e.g., in middle and high school, in co-teaching environments, or for special education teachers pushing into general education classrooms). The models in this paper use data from the survey administered to fourth graders in 2015-16 as the pretest measure of students' SEL, and data from the survey administered to those same students as fifth graders in 2016-17 as the posttest measure of students' SEL.

We limit our analysis dataset to those fifth-grade students who are linked to one and only one teacher for instruction in all observed subjects. In some cases, this means that a student is linked to a fifth-grade "homeroom teacher;" in other cases, this means that students are attributed to the same fifth-grade teacher for instruction in both math and ELA. Of the more than 40,000 students in the sample, we exclude 3,225 students from the analysis who are linked to more than one teacher for a subject. We exclude 3,985 students from the analysis who have different teachers for math and ELA.

Figure 1 displays the number of fifth-grade teachers per school in the sample (limited to teachers linked to at least 10 students). When there is more than one fifth-grade teacher in a given school, we can disentangle across-school differences in students' SEL from across-

teacher-within-school differences. The number of fifth-grade teachers per school in our sample ranges from one to 11, and the modal number of fifth-grade teachers per school is three.

Figure 1. Number of Fifth-Grade Teachers per School

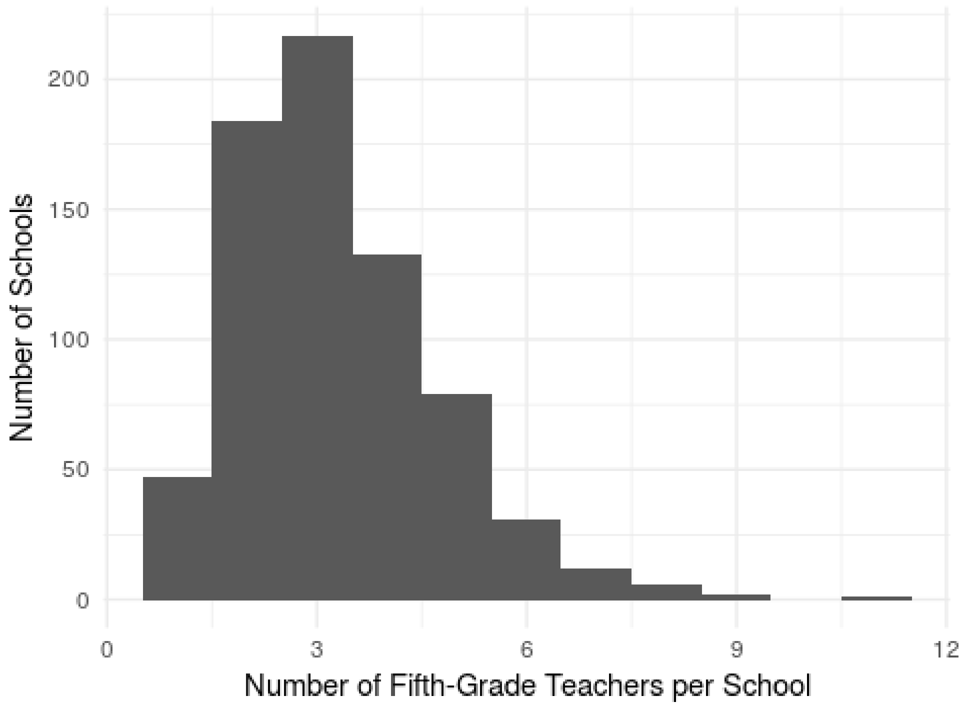
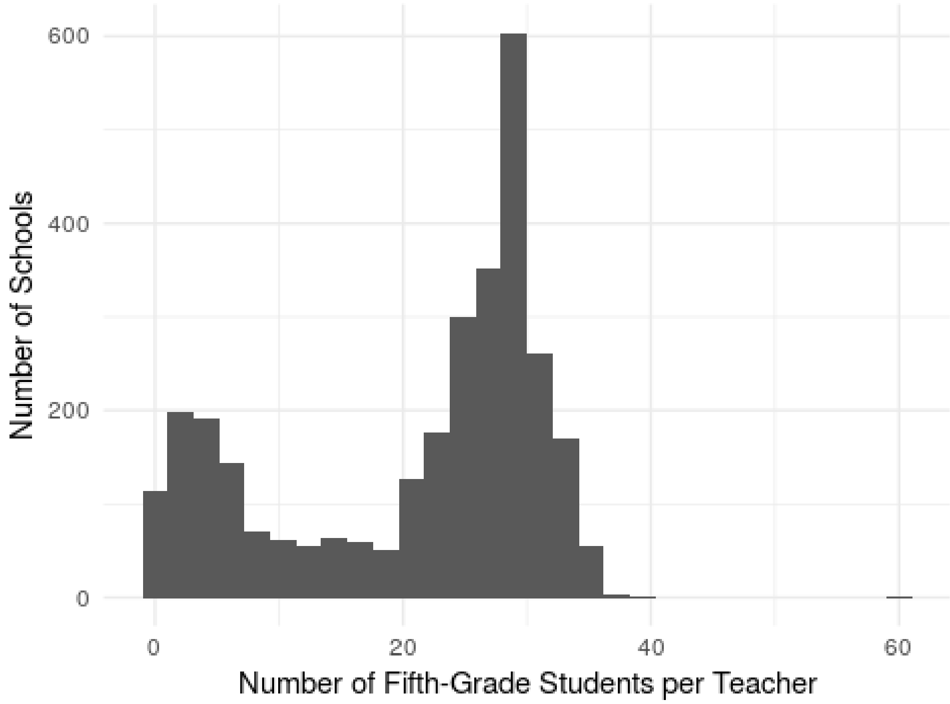


Figure 2 displays the distribution of the number of students per teacher in our dataset. We have defined the sample to include teachers who appear (via student-teacher linkage) to be responsible for primary instruction of both math and ELA. As we would expect, most of these teachers are linked to between 20 and 40 students.

Figure 2. Number of Fifth-Grade Students Per Teacher



In addition to the student SEL survey data, we use data from the Smarter Balanced Assessment Consortium (SBAC) math and ELA assessments, which are computer-adaptive assessments aligned to the Common Core standards administered to students in grades three through eight across California. Students completed these assessments in the spring of 2015-16 and 2016-17, which enables us to compare students' growth in SEL to their growth in math and ELA achievement.

The samples we use in producing the SEL growth measures comprise students in the CORE Districts who responded to the survey in both 2015-16 and 2016-17. For their responses to be considered valid for inclusion in the sample, students need to have responded to at least half of the survey items within a construct. In order to be included in the growth measure for a given SEL construct, students must have had valid survey responses in 2016-17 for that particular construct (i.e., the posttest), as well as valid responses in 2015-16 for all four constructs (i.e., the pretest measures). We required valid responses to all four constructs in 2015-16 because all four are control variables in the growth model. In addition, students must have had SBAC scores in math and ELA in 2015-16 and demographic data available to serve as additional control variables in the growth model. Similarly, we estimate the SBAC growth measures for a given subject using a sample of students in CORE with SBAC scores in that subject in 2016-17, SBAC scores in both subjects in 2015-16, valid responses in all four SEL constructs in 2015-16, and available demographic data.

Table 1 below summarizes the demographic characteristics of the fifth-grade students in the sample. Note that depending on the outcome of interest (math, ELA, or the four SEL constructs), the sample differs slightly depending on whether they completed the given outcome assessment. Averaging across the six outcome measures (n_{avg} = 43,576), 16.9 percent of students are English Language Learners (ELL), 10.4 percent are students with disabilities (SWD), 79.2 percent are economically disadvantaged, 3.7 percent are homeless, 0.44 percent are in foster care, 71.6 percent are Latinx, 10.0 percent are White, 7.1 percent are Asian, and 7.0 percent are African American.

Table 1. Characteristics of Students in the Sample

Outcome	N of Students	% ELL	% SWD	% Econ Disadv	% Homeless	% Foster	% Latinx	% White	% African American	% Asian
Math	46408	17.1	10.75	79.15	3.65	0.44	71.51	10.07	7.1	6.98
ELA	46392	17.1	10.77	79.14	3.65	0.44	71.5	10.09	7.1	6.98
Growth Mindset	41600	16.56	10.16	79.18	3.7	0.44	71.54	9.98	6.88	7.2
Self-Efficacy	41575	16.54	10.16	79.18	3.7	0.44	71.55	9.97	6.87	7.2
Self-Management	42000	16.62	10.24	79.15	3.69	0.44	71.53	9.99	6.91	7.17
Social Awareness	42136	16.67	10.28	79.17	3.69	0.44	71.55	9.99	6.92	7.15

Methods

Before describing the methods to address each of the five research questions, we first establish some notation. We observe measures of students' SEL over two school years, 2015-16 and 2016-17, and in four constructs: Growth Mindset, Self-Efficacy, Self-Management, and Social Awareness. We also include standardized test scores in Math and ELA in all analyses, providing a total of six measured outcomes, or "constructs." Each of these six constructs is represented by c.

Let Y_{cikjt} be the scale score for construct c for student i associated with classroom k in school j in year t. This is the scale score for a given student in construct c in 2016-17. Note that the student is associated with only one classroom and only one school in that year.

Let $Y_{cikjt-1}$ be the scale score for the same student and the same construct in the prior year, 2015-16. We do not make any assumptions about what school the student attended or what classroom they were associated with in the prior year.

RQ1. Are There Classroom-Level Differences in Students' SEL?

Our first research question investigates whether classrooms differ with respect to SEL outcome measures. One hypothesis might be that all variation in SEL outcome measures is random with respect to classrooms, such that one is equally likely to find low or high SEL in any

classroom (or alternatively, that observed low or high SEL in a classroom is a result of random error). To investigate this, we estimate a model in which we predict students' SEL outcome measures with a random effect for which classroom they were associated with in 2017 (in other words, which classroom they were in). The estimate of the variance of this random classroom effect allows us to determine whether there are systematic classroom differences in mean SEL outcome measures.

Because classrooms are nested within schools, we also need to account for the part of the variance that may be attributable to the school; this is important so that we avoid over-interpreting classroom-level differences that may actually be driven by school differences. This means we include nested random effects for schools and classrooms to predict the outcome in the estimation of this model.

The model decomposes SEL into parts that are due to the school the student attends, classroom the student is in, and some additional unexplained student component. We can write this as:

$$\boldsymbol{Y_{cikjt} = \gamma_{cjt} + \mu_{ckjt} + \eta_{cikjt}} \tag{1}$$

Where:

- Y_{cikjt} is the construct outcome for students' math, ELA, or SEL outcome measure in 2017,
- γ_{cjt} is the component of the variance that is across schools
- μ_{ckjt} is the component of the variance that is within-schools-across-classrooms, and
- η_{cikjt} is the component of the variance that is within classrooms.

RQ2. Are There Differences in Students' Prior Levels of SEL?

To answer our second research question, we estimate the same model specified in Equation (1) above, except with the SEL measures in year $t-1$ as the outcome variable. These models still include classroom k and school j in year t. This can be interpreted as quantifying the degree of differences in students' prior levels of SEL (or their "starting point") by classroom and by school.

RQ3. Can We Detect Classroom-Level Effects of Students' Growth in SEL?

Our third research question asks whether there are differences in students' growth in SEL from 2016 to 2017 that can be attributed to the classroom. To answer this, we construct a model of student growth in which we predict a student's outcome (i.e., the student's scale score for math, ELA, or one of the four SEL constructs) with lagged scores of all six scale scores in the prior year, student demographic characteristics (ELL, SWD, economic disadvantage, homelessness, foster care, race/ethnicity), and a variable for the classroom the student was assigned to. This model can be written as:

$$\boldsymbol{Y_{cikjt} = \xi_c + Y_{cikjt-1}\lambda_c + X_{ikjt}\beta_c + \alpha_{ckjt} + e_{cikjt}} \tag{2}$$

Where:

- Y_{cikjt} is the construct outcome for a student (Math, ELA, or SEL scale score) in 2016-17,
- $Y_{cijkt-1}$ is a 1x6 vector of lagged outcome measures (Math, ELA, or SEL scale score) for student i in year $t-1$,
- X_{ikjt} is a vector of characteristics for student i in year t,
- α_{ckjt} is the impact of classroom k in school j on growth in construct c in year t, and
- e_{cikjt} is a student-level error term.

We estimate this regression using an errors-in-variables (EIV) method described by Fuller (1987) that accounts for measurement error in $Y_{cijkt-1}$. We use estimates of Cronbach's alpha for lagged SEL constructs, and conditional standard errors of measurement for SBAC scores (note that because the SBAC is a computer-adaptive test, Cronbach's alpha does not apply).

RQ4. How "Big" or "Small" are Classroom-Level Effects of Students' Growth in SEL?

Our fourth research question asks how we can interpret the magnitude of any classroom-level effects of student growth in SEL identified in the previous question. To estimate the magnitude of classroom-level effects, we directly estimate the parameter α_{ckjt} in the EIV regression in Equation (2) and compute an estimate of the variance of the classroom effect α_{ckjt} using Equation (3):

$$\boldsymbol{\widehat{V}\,(\alpha_{ckjt}) = V(\widehat{\alpha}_{ckjt}) - \overline{\sigma}^2_{ckjt}} \tag{3}$$

Where:

- $V(\widehat{\alpha}_{ckjt})$ is the sample variance of the classroom effect, and
- $\overline{\sigma}^2_{ckjt}$ is the mean squared standard error of the classroom effect.

This variance is a measure of the magnitude of classroom effects. If we find little variance in classroom effects, this suggests there are not large differences in growth across classrooms.

To further study the degree to which student SEL growth is attributable to schools and classrooms, we attempt to decompose growth into components attributable to school, classroom, and student. To do so, we construct a student-level estimate of growth, and then apply the variance decomposition framework from Equation (1). We need to construct a student-level estimate of growth by computing a residual based on a prediction that excludes the parameter α_{ckjt}. This residual contains α_{ckjt} plus the random student error e_{cikjt}, and as

such, an average of these student-level growth measures for a particular classroom k in construct c is equal to the estimate of α_{ckjt}. We can write this residual as:

$$\boldsymbol{\epsilon_{cikjt}} = \boldsymbol{Y_{cikjt}} - \boldsymbol{Y_{cikjt-1}}\,\widehat{\boldsymbol{\lambda_c}} - \boldsymbol{X_{ikjt}}\,\widehat{\boldsymbol{\beta_c}} \tag{4}$$

We then use the same model setup for variance decomposition to find out how much variance of student-level growth is explained by the student's school and the student's classroom:

$$\boldsymbol{\epsilon_{cikjt}} = \boldsymbol{\gamma_{cjt}} + \boldsymbol{\mu_{ckjt}} + \boldsymbol{\eta_{cikjt}} \tag{5}$$

The sum of the estimates of the variances of school and classroom in Equation (5) are another way to estimate $V(\alpha_{ckjt})$ obtained from the growth model in Equation (2). However, setting up the model in this way allows us to investigate the contributions of school and classroom effects to estimates of growth separately.

RQ5. Do Classrooms with High SEL Growth Also Have High Growth in Academic Outcomes?

Our final research question aims to examine whether the classrooms that experience high growth in the four SEL constructs also show high growth in academic outcomes (i.e., math and ELA). To answer this question, we use estimates of α_{ckjt} to compute correlations between classroom effects among the six outcome measures.

Results

RQ1. Are There Classroom-Level Differences in Students' SEL?

An important prerequisite to estimating classroom value-added models of SEL outcomes is to determine whether the SEL outcomes *themselves* differ across classrooms. In previous work (Loeb et al., 2019, Fricke et al., 2019), we examined the proportion of variance in students' SEL attributable to the school a student attends. Using the student-teacher linkage data in the current paper, we expand upon this prior work to further decompose the variance in students' SEL at the classroom level. If we were to find that classroom-level outcomes do not differ more than would be expected by randomness in student growth, this would suggest that there are not classroom-level effects on students' SEL. However, if we were to find that classrooms do differ overall, but that classrooms within the same schools tend to be similar, this would suggest that differences in students' SEL may be attributable only to school-level differences. Finally, if we were to find that even classrooms within schools show differences in students' SEL outcomes, this would motivate us to examine whether classrooms also differ in students' growth in SEL from one year to the next.

To investigate these various possibilities, we apply the variance decomposition framework described earlier to estimate the variance (i) across schools, (ii) across classrooms and within schools, and (iii) within classrooms (i.e., student-level differences). We include math and ELA outcomes in this analysis in order to compare the magnitude of classroom differences in SEL to those seen in math and ELA standardized assessments in this sample.

Table 2 below shows this variance decomposition as the proportion of variance explained at each of the three levels. We see that, similar to math and ELA outcomes, the magnitude of the differences across-classroom-within-school is larger for each outcome (0.37 for math, 0.38 for ELA, 0.14 for growth mindset, 0.04 for self-efficacy, 0.05 for self-management, and 0.05 for social awareness) than the magnitude of differences across-school (respectively, 0.12, 0.09, 0.06, 0.03, 0.02, and 0.03). However, a much smaller proportion of the variance is due to across-classroom-within-school differences in SEL compared to across-classroom-within-school differences in math or ELA. Notably, growth mindset stands out as having comparatively large across-classroom-within-school differences relative to the other SEL constructs. As we note in the discussion, there are some undesirable measurement properties of the growth mindset construct that may explain this distinction. Taken together, these results suggest that (i) a greater proportion of variance in students' SEL outcomes is explained at the across-classroom-within-school level than the across-school level for each outcome, (ii) we are more likely to see a very wide range of SEL outcomes within classrooms in a given school than we are to see a very wide range of math or ELA outcomes, and (iii) growth mindset has a higher proportion of variance explained at the across-classroom-within-school level than the other SEL constructs.

Table 2. Variance Decomposition of Grade 5 Models: Proportion of Variance of Posttest as Outcome

Outcome	Across-School	Across-Classroom-Within-School	Within-Classroom
Math	0.12	0.37	0.51
ELA	0.09	0.38	0.53
Growth Mindset	0.06	0.14	0.80
Self-Efficacy	0.03	0.04	0.93
Self-Management	0.02	0.05	0.93
Social Awareness	0.03	0.05	0.92

RQ2. Are There Differences in Students' Prior Levels of SEL?

We repeat this variance decomposition analysis for the scale scores of students in a classroom from the prior year, regardless of which school they attended, in order to quantify differences in students' "starting points." If all students were randomly assigned to schools and to classrooms in the sample each year, then we would expect no variance in prior year scores to be attributable to across-school or across-classroom-within-school differences.

However, as Table 3 below suggests, this is not the case in our sample. Comparing Table 2 with Table 3 shows similar proportions of variance in prior outcomes in year $t-1$ (as shown in Table 3) to the proportions of variance in current outcomes in year t (as shown in Table 2); this suggests there are differences in students' prior SEL and prior academic achievement at both the school and classroom levels. Specifically, as we have shown in prior work (Loeb et al., 2019; Fricke et al. 2019), the across-school variance in prior SEL is smaller than the across-school variance in math and ELA; this suggests different schools serve students with different prior math and ELA achievement and to a lesser degree, with different prior SEL. The results in Table 3 extend this by showing that the across-classroom-within-school differences in prior SEL are smaller than across-classroom-within-school differences in prior academic achievement; nevertheless, the results indicate that classrooms within the schools in our sample are serving students with different starting points in achievement and to a lesser degree, in SEL. Consequently, this motivates the need to take these differing starting points into account when attempting to measure how much of an impact a student's classroom has on their growth in SEL over the course of a school year.

Table 3. Variance Decomposition of Grade 5 Models: Proportion of Variance of Pretest as Outcome

Outcome	Across-School	Across-Classroom-Within-School	Within-Classroom
Math	0.13	0.37	0.51
ELA	0.12	0.35	0.53
Growth Mindset	0.08	0.09	0.83
Self-Efficacy	0.03	0.06	0.91
Self-Management	0.04	0.07	0.89
Social Awareness	0.03	0.03	0.94

RQ3. Can We Detect Classroom-Level Effects of Students' Growth in SEL?

Given that students differ in terms of their prior year SEL within classrooms, we estimated a value-added model from Equation (2) above, with a goal of retrieving classroom-level effects on students' growth in SEL. We report some diagnostics from these models below, but we mostly focus on measures of the variance of the classroom effects, because the detailed properties of these models have been discussed in prior work (Loeb et al., 2019). The estimation of the current models differs from those in Loeb et al. (2019) and Fricke et al. (2019) only in that we have included fixed effects for classrooms, rather than for schools.

To summarize the predictive power of the models, Table 4 below reports a goodness-of-fit measure that summarizes the predictive power of the regression model *excluding* the predictive power of fixed effects for the classrooms (i.e., within-classroom R^2 based on within-classroom variance only). The goal of this measure is to report only the predictive power of prior score outcomes and other student characteristics. A value-added model with high predictive power more strongly suggests that differences in student characteristics are

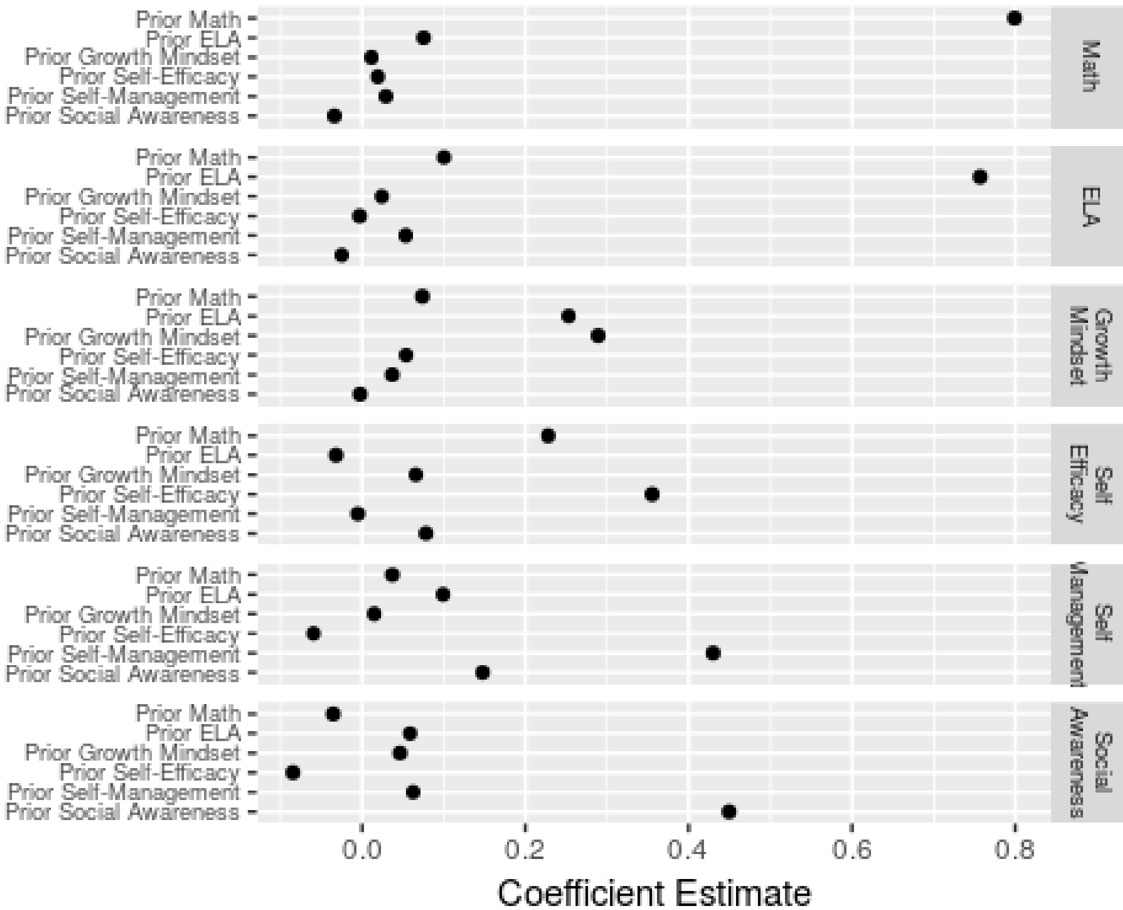
sufficiently controlled for. As noted in Loeb et al. (2019), the R^2 is quite low for the SEL constructs compared to typical value-added models of academic achievement. This presents a challenge for the interpretation of the classroom effects, and thus warrants general caution against overinterpreting the results from these models.

Table 4. Goodness-of-Fit for Grade 5 Classroom Growth Models

Outcome	Within-Classroom R^2
Math	0.70
ELA	0.68
Growth Mindset	0.19
Self-Efficacy	0.20
Self-Management	0.24
Social Awareness	0.16

Figure 3 below displays the coefficient of each control variable on each of the six outcomes. The grey boxes on the right indicate the outcome measure of the particular value-added model. Each dot represents a coefficient in that model for the control variable indicated on the left side of the figure. As would be expected, the largest coefficient for each model is the prior year measure of the same outcome. In addition, the prior SEL measures are less predictive of the current year SEL measures than the prior ELA or Math measures are of the current year ELA or Math measures (respectively), as evident by the smaller prior-year same-construct coefficients in each SEL model relative to the prior-year same-subject coefficients in the academic models.

Figure 3. Coefficient Estimates from Each Model



RQ4. How "Big" or "Small" are Classroom-Level Effects of Students' Growth in SEL?

Given the models estimated in the previous section, we next examine the distribution of the classroom effects α_{ckjt} to determine the extent to which we can identify differences in classroom impacts on students' growth in SEL. We estimate the variance of the classroom effects α_{ckjt} by using Equation (3), which characterizes the extent to which classroom effects differ, in standardized units of the outcome measure. The following table displays the error-corrected standard deviation of the classroom effects.

Table 5. Error-Corrected Standard Deviation of Classroom Effects

Outcome	True Standard Deviation
Math	0.25
ELA	0.19
Growth Mindset	0.31
Self-Efficacy	0.27
Self-Management	0.22
Social Awareness	0.26

The error-corrected standard deviation of classroom effects is a summary statistic for the distribution of classroom effects displayed in Figure 4 below, which shows in rank order the point estimates of the classroom effects with 95% confidence intervals. Compared to the growth measures of ELA and math, the SEL growth measures tend to have more variance (as indicated by the range of the highest and lowest points of the red line, i.e., the classroom effects) but higher standard errors (as indicated by the black lines, i.e., the confidence intervals). Figure 5 below presents the distribution of the point estimates themselves; as the figure indicates, the distribution of the classroom effects is approximately normal.

Figure 4. Fifth-Grade Classroom Estimates by Construct

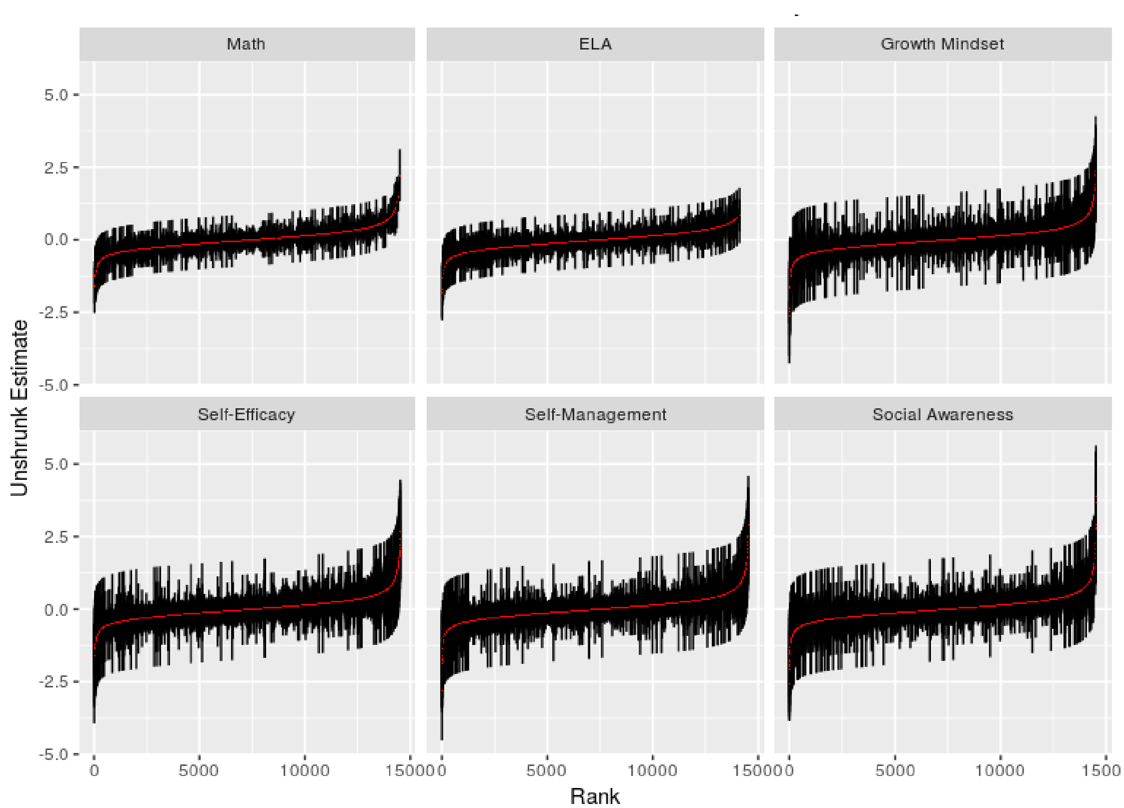
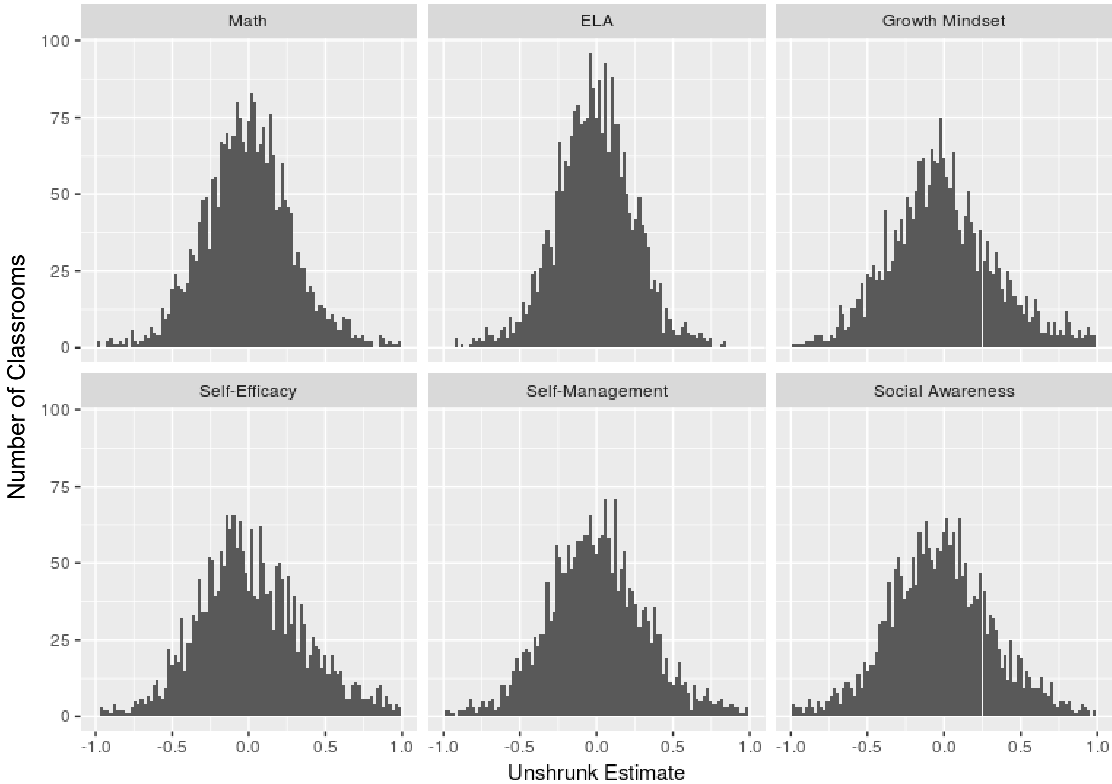


Figure 5. Fifth-Grade Classroom Estimates by Construct



The overall differences among classroom-level effects indicate whether classrooms have different impacts on students' growth in SEL. However, we are also interested in assessing the degree to which those classroom impact differences are driven by school-level impacts. If some schools tend to have many high-growth classrooms and other schools many low-growth classrooms, the overall differences in classrooms described in the overall variance of α_{ckjt} still exist, but can be partially explained by school differences.

To investigate the differences in growth that we can attribute to classrooms rather than to schools or to random student variance, we construct a student-level measure of growth by retrieving a student residual from the value-added model. This residual is the amount by which the student exceeded or fell short of their predicted score, based on their prior year scores and other student characteristics. We then use the variance decomposition framework described earlier on the residual to estimate the variance of the school and classroom components of student growth. In doing so, we attempt to explain student-level differences in growth with the school they attended and the classroom they were assigned to.

In Table 6 below, we report the variance of student growth attributable to classrooms and schools. The student-level residual contains both the classroom effect α_{ckjt} as well as the random error e_{cikjt}. The sum of the variances in each row of the table is the total variance of the residuals in the growth model. Due to the lower predictive power of the models with SEL

measures as outcomes, the variance of the residuals in the models (i.e., the sum of each row) is much higher for the SEL outcomes than for the math and ELA models. In other words, outcomes are less consistent from year to year in the SEL measures than in the academic measures, so there is more student-level "growth" to explain. As a result, the variance of residuals attributable to classrooms (i.e., the third column from the left in Table 6) is similar in *magnitude* for all six outcomes, but it is a smaller *proportion* of the total variance than in Math or ELA. Table 7 more clearly conveys this, by reporting the *percentage* of the total variance in growth explained by classrooms.

Table 6. Variance Decomposition of Grade 5 Classroom Growth Models: Growth as Outcome

Outcome	Across-School	Across-Classroom-Within-School	Within-Classroom
Math	0.02	0.05	0.21
ELA	0.01	0.03	0.24
Growth Mindset	0.02	0.07	0.69
Self-Efficacy	0.02	0.05	0.77
Self-Management	0.01	0.04	0.74
Social Awareness	0.02	0.05	0.82

Table 7. Variance Decomposition of Grade 5 Classroom Growth Models, Shown as Percents: Growth as Outcome

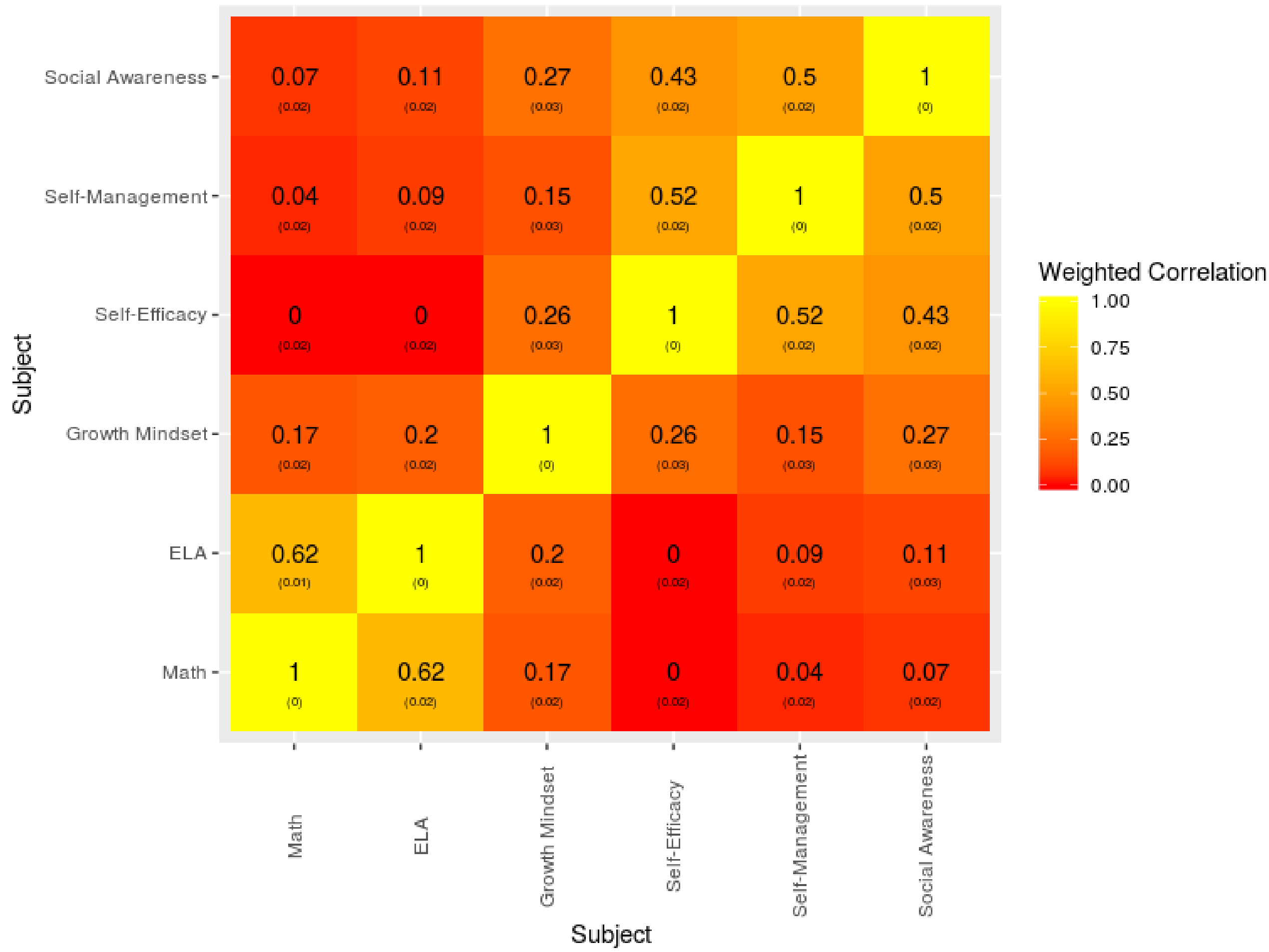
Outcome	Across-School	Across-Classroom-Within-School	Within-Classroom
Math	7%	17%	77%
ELA	4%	10%	86%
Growth Mindset	3%	9%	88%
Self-Efficacy	2%	6%	92%
Self-Management	1%	5%	94%
Social Awareness	2%	5%	93%

The sum of the variances of the across-school and across-classroom-within-school components is equivalent to an estimate of the variance of the classroom fixed effects α_{ckjt}. We can think of this sum as a "combined" estimate of the magnitude of the across-classroom effect on student growth (i.e., including both classroom and school impacts), because this variance describes the distribution of classroom impacts.

RQ5. Do Classrooms with High SEL Growth Also Have High Growth in Academic Outcomes?

Our final research question asks whether classrooms showing high growth in SEL also show high growth in academic outcomes. To answer this, we examine the associations between classroom effects on the six different outcomes to see if classrooms with impacts in some outcomes tend to have similar impacts on other outcomes. Figure 6 below displays correlations between the classroom effects for each of the six outcomes.

Figure 6. Weighted correlations (and standard errors) among classroom effects on each outcome



The strongest relationships are between growth in the two academic subjects, math and ELA ($r = 0.62$). Growth in social awareness, self-management, and self-efficacy are all strongly related to one another as well ($0.43 \leq r \leq 0.52$). Growth mindset again stands apart from other SEL constructs, because it is the SEL construct most strongly correlated with math and ELA, and is the most weakly correlated with the other three SEL constructs; as we note in the discussion below, this is potentially due to some undesirable measurement properties of the growth mindset construct. Taken together, these results indicate that classrooms having a large impact on students' growth in math and ELA are not necessarily the same classrooms having a large impact on their growth in SEL (or vice versa); however, there are some classrooms having a large impact on students' growth mindset that also appear to have an impact on students' academic achievement.

Discussion

The results in this paper provide preliminary insights into the degree to which we can measure classroom impacts on students' self-reported SEL, and thus builds upon an emerging body of research aimed at establishing the magnitude of classroom-specific impacts on students' non-cognitive growth and development (e.g., Blazar, 2018; Blazar & Kraft, 2017; Jackson, 2018; Jennings & DiPrete, 2010; Ruzek et al., 2015). We discuss the findings from each of our five research questions in the sections that follow, and then present a more general discussion.

RQ1. Are There Classroom-Level Differences in Students' SEL?

We first aimed to parse out whether there are classroom-level differences in students' SEL outcomes, and how much of those differences can be attributed to classrooms specifically, rather than schools more broadly. Our results showed that there are indeed classroom-level differences in students' SEL outcomes, even after accounting for school-level variance; however, classroom differences are smaller for the four SEL outcomes measured than for academic outcomes (i.e., math and ELA).

Our prior work (Loeb et al., 2019) established this same pattern for school-level variance, but the current paper builds upon these findings by establishing that there is an additional source of measurable variance—at the across-classroom-within-school level. We found that across-classroom variance is slightly larger, in fact, than the across-school variance, particularly for the growth mindset outcome. We know from prior work that the growth mindset construct has lower internal consistency as measured by Cronbach's alpha than the other SEL constructs in the CORE survey (Meyer, Wang & Rice, 2018). Our prior work also found that students—especially those in earlier grades—were more likely to exhibit rating-scale confusion on the growth mindset construct compared to the other SEL constructs (Bolt, Wang, Meyer, & Rice, 2018), potentially due to the negative wording of that construct; this confusion was correlated with students' age, reading proficiency, and ELL status—suggesting that the growth mindset construct might be measuring English reading ability more than the other constructs. Therefore, although additional research is needed to conclude what it is about the growth mindset measure specifically that might have greater variance across classrooms, the findings here align with our prior work showing there may be something unique (and perhaps undesirably so) about the growth mindset construct in particular that is worthy of further investigation.

RQ2. Are There Differences in Students' Prior Levels of SEL?

We next explored whether classrooms differ in students' *prior* levels of SEL. We found similar patterns in across-classroom differences of students' prior SEL (i.e., pretests) as we did in their SEL outcome measures (i.e., posttests): Across-classroom differences were smaller for the SEL pretest measures than the academic pretest measures, and there was slightly larger across-classroom-within-school variance than across-school variance. This indicates that

students within a particular classroom are similar to each other in terms of their prior academic performance (i.e., evidence of some degree of "tracking") and, to a lesser extent, their prior growth mindset, self-efficacy, self-management, and social awareness. Consequently, it is important to account for these classroom-level pretest differences if we want to truly answer the question of whether different classrooms might differentially support students' growth in SEL over time. Thus, we next turn to examining whether we can use these pretest and posttest measures in growth models in pursuit of this goal.

RQ3. Can We Detect Classroom-Level Effects of Students' Growth in SEL?

We examined whether we can detect these classroom-level impacts on students' growth in SEL. Although we were able to detect an effect across classrooms, the lower explanatory power (i.e., within-classroom R^2) of the SEL models relative to the academic models means that it is less clear that these effects are causal effects that have appropriately controlled for students' starting points. Even if we *are* measuring causal effects, the lower explanatory power of the SEL measures calls into question the long-term relevance of those impacts. However, the results in this paper nonetheless extend our prior work applying school-level value-added models to these SEL measures by establishing that we can account for additional SEL growth by knowing what fifth-grade classroom a student was in, above and beyond the school in which he or she was enrolled. This is important for establishing that the results of the school-level growth models were not simply indicative of school-level error (for example, whether it was a rainy day in the first year and a sunny day in the second year, so most students in a school were likely to respond more positively in year two than in year one). This does not rule out the possibility of classroom-level error, as well (for example, whether a classroom of students had a difficult quiz the day before the SEL survey in one year but watched a movie in class the second year), but does suggest there is signal and not just noise. In addition, if there were more measurement error in the SEL survey than the current approach (i.e., the errors-in-variables regression) implies, this would further limit the interpretation of the results from the SEL growth models. Thus, we recommend interpreting the results presented here with caution—in particular given the low within-classroom R^2 of the SEL models; nonetheless, we believe these results indicate there might be measurable student growth in students' responses to this SEL survey that is impacted in some way by the environment of the classroom *and* the school.

RQ4. How "Big" or "Small" are Classroom-Level Effects of Students' Growth in SEL?

Our fourth research question aimed to assess the magnitude of the classroom-level effects we detected. The results showed that more student growth in SEL is explained by classroom differences within a school than differences across schools. A classroom that produces student growth in SEL that is one standard deviation above average produces similar gains (in error-corrected standard deviation units of the outcome measures) as a classroom shows for math or ELA (ranging from 0.22 to 0.31 standard deviations for the SEL constructs, compared to 0.19 for ELA and 0.25 for math). In other words, we can think of the classroom-level impact on students' growth in SEL as similar to those impacts on students' growth in

academic outcomes. However, there is *proportionally* less variance explained at the classroom level in SEL than academics, because the within-classroom (across-student) variance is much higher in SEL. Another way of saying this is that the magnitude of the impact is similar in SEL and in academic outcomes, but the higher variability year-to-year of SEL compared to academics means there is proportionally less variance explained by classrooms for SEL compared to academics (and proportionally more variance explained at the individual student level).

RQ5. Do Classrooms with High SEL Growth Also Have High Academic Growth?

Our final research question probed whether the same classrooms producing high growth in SEL also produce high growth in academic outcomes. We found that this is not generally the case. Results showed that the correlation between growth in terms of SEL and growth in math or ELA is close to zero—except for the growth mindset construct. In fact, we found that classroom-level growth in growth mindset is more correlated with growth in ELA than with growth in the three other SEL constructs. Again, this aligns with our prior work showing that students who were younger, who were ELLs, or who had lower scores on their ELA assessments were more likely to exhibit rating-scale confusion on the growth mindset construct (Bolt et al., 2018), suggesting the growth mindset items are a measure, in part, of a student's English reading proficiency. In general, the results presented here indicate that the SEL growth measures capture classroom-level supports that are not captured by growth in academic test scores alone; we believe such impacts may be important to consider as the field weighs various options for expanding the definition of student success in school.

General Discussion

In this paper, we estimated standard deviations ranging from 0.10-0.14 for the SEL measures at the school level, 0.26-0.30 at the school-plus-classroom level, and 0.20-0.26 at the classroom level after accounting for school-level effects. Overall, the results in this paper align with findings from recent studies aimed at quantifying classroom-level effects on students' non-cognitive measures. Blazar and Kraft (2017) measured classroom impacts of similar magnitudes for students' self-efficacy in math (0.14 standard deviations) and for math academic performance (0.18 standard deviations) as we found in this paper (i.e., 0.25 standard deviations for math, 0.19 standard deviations for ELA, and standard deviations ranging from 0.22 to 0.31 for the four SEL constructs). Blazar (2018) not only established that the magnitude of the classroom-level effects for non-cognitive measures was similar to (or larger than) effects on math test scores, but also found causal effects for these variables. Ruzek and colleagues (2015) found that classroom effects on students' mastery goal orientation was roughly equivalent in magnitude to effects on students' math achievement, which also aligns with the findings in this paper. Finally, Jennings and DiPrete (2010) established an effect size of 0.28 at the school level for social/behavioral measures, of 0.35 at the school-plus-teacher level, and of 0.21 at the classroom level after accounting for school-level effects. Our findings differ in that we found more variance explained at the across-classroom-within-school level than at the school level.

Jackson (2018) reported classroom impacts on students' non-cognitive skills that were 10 times more predictive of long-term success in high school than impacts on academic test scores. We do not currently have the data to attempt to replicate such predictions of long-term outcomes from estimates of fifth-grade classroom impacts; in addition, we make use of data that do not require us to make assumptions or restrictions about the impacts of other teachers who are also assigned to the same students. Although Jackson examines differences within a particular track (i.e., advanced versus standard rigor) within a school, the results here document that there are differences in outcomes across classrooms within schools more specifically.

Taken together, the results presented here generally reflect what others have found using other non-cognitive outcome measures: there seem to be classroom-level effects, above and beyond school-level effects, on students' growth in non-cognitive measures that could be on par with impacts on students' academic growth. A critical area for future research is to assess the degree to which the SEL growth identified here persists from year to year; given that the predictive power of the SEL models are lower than academic growth models, it is unclear whether we can measure the long-lasting impacts of being in a classroom that produces high SEL growth. Although Jackson (2018) suggests this may be the case for predicting other long-term outcomes, additional data are needed to examine the degree to which SEL growth produced by classrooms is a persistent effect over time. This is one of many interesting and important avenues for further investigation.

Limitations

There are three important caveats to the findings presented here. First, the validity of the classroom-level SEL growth measures estimated depends upon the validity of the SEL survey items that underlie them. Importantly, surveys are not the only way to measure SEL outcomes; outcomes such as suspension rates or chronic absenteeism (which are currently included in CORE's school performance dashboard in addition to the SEL survey measures) can also be used as proxies for SEL. These alternative measures have the benefit of not relying on student self-reports; however, they also have the disadvantage of not clearly distinguishing the particular SEL skills of interest, and are known to be confounded with other variables separate from SEL, such as race/ethnicity or socioeconomic status.

Second, an additional limitation of the current CORE SEL survey is that it is administered once annually. If students' SEL is not stable and changes over time, then students' responses may change depending on when the survey is administered (i.e., be subject to "occasion error"). Consequently, measuring SEL once a year will not capture the "true," persistent component of students' SEL. Future experimental research could leverage a test-retest approach in which surveys are administered at least twice in a relatively short time span (e.g., two to four weeks apart) in order to assess the extent to which the surveys are subject to occasion error.

Finally, comparing SEL growth models to their academic counterparts does not necessarily establish the validity of the SEL models as a measure of classroom effects on SEL outcomes. In addition, given the relative newness of the CORE survey measures and potential measurement issues identified with the constructs (Bolt, et al., 2018; Meyer et al., 2018), we strongly caution against interpreting the growth measures reported here as causal estimates of classroom effects on students' SEL. However, classroom-level SEL growth measures can be useful even when it is not clear the extent to which the measures are causal. The results in this paper suggest there may be measurable differences in SEL growth from one classroom to the next within a given school; with additional research to probe the robustness of this finding, this could help educators and administrators identify effective classroom supports that are positively impacting students' non-cognitive development beyond their expected trajectory. Similarly, such measures could help educators identify classrooms in which students may be falling behind their peers in order to deploy additional resources, such as referrals to school-based clinicians, additional teacher professional development opportunities, or more intensive interventions through Multi-Tiered Systems of Support (MTSS) or Response to Intervention (RTI) frameworks in place within a school. However, before any such measures could be considered for use in any sort of intervention, evaluation, or accountability system, additional research establishing the validity of the measures for such uses is critical.

Conclusion

Using data from a large-scale SEL survey of more than 40,000 fifth-grade students across five large districts in California, we produced and evaluated measures of the impacts of individual classrooms on students' SEL outcomes. In doing so, we build upon an emerging but growing body of empirical research assessing the degree to which classrooms impact students' SEL, above and beyond the effects that a school's culture or climate may have. We produced value-added models for four SEL constructs (growth mindset, self-efficacy, self-management, and social awareness) by applying methodology and model specifications similar to those often used to measure the impacts of classrooms on academic subjects, such as math and ELA, which we also estimated for classrooms in the districts administering the survey.

We identified across-classroom-within-school variance of students' SEL outcomes, even after accounting for school-level variance. We found magnitudes of classroom-level impacts on students' growth in SEL similar to impacts on students' growth in ELA and math, but we showed that growth models of SEL do not perform as well as growth models of academic outcomes. Importantly, results indicated that across-classroom-within-school impacts were larger in magnitude than across-school impacts on students' SEL growth. Finally, we found low correlations between classroom-level growth in SEL and classroom-level growth in ELA or math; however, growth mindset stands apart from the other three SEL constructs in that there is a moderately strong relationship—one which may be a function of its poorer psychometric properties.

As conversations around appropriate and valid uses of measures of students' SEL continue—particularly as states make further progress to specify the "fifth indicator" of their

ESSA plans, and districts in California develop and report local measures of school climate—this paper provides some preliminary evidence as to whether we can reliably measure classroom-level impacts on students' non-cognitive skills using the CORE self-report SEL surveys. In doing so, we aim to contribute to the growing body of knowledge about appropriate and innovative uses of measures of students' non-cognitive skills and social-emotional learning (McKown & Taylor, 2018). In addition, these results set the stage for additional research investigating how and whether classroom-level supports might causally produce measurable impacts on students' SEL, which further informs how schools and districts think about programs, policies, and initiatives to improve students' SEL.

References

American Institutes for Research (AIR). (2017). *Examining ESSA plans through the lens of research and practice: Reflections on state ESSA plans*. Washington, DC: American Institutes for Research.

Bacher-Hicks, A., Chin, M., Kane, T. K., & Staiger, D. O. (2017). *An evaluation of bias in three measures of teacher quality: Value-added, classroom observations, and student surveys (NBER Working Paper No. 23478).* Cambridge, MA: National Bureau of Economic Research.

Berg, J., Osher, D., Same, M. R., Nolan, E., Benson, D., & Jacobs, N. (2017). *Identifying, defining, and measuring social and emotional competencies: Final report.* Washington, DC: American Institutes for Research.

Betts, J. R., & Tang, Y. E. (2008). Value-added and experimental studies of the effect of charter schools on student achievement: A literature review. Seattle, WA: National Charter School Research Project, Center on Reinventing Public Education.

Biancarosa, G., Bryk, A. S., & Dexter, E. R. (2010). Assessing the value-added effects of literacy collaborative professional development on student learning. *Elementary School Journal, 111*(1), 7-34.

Blazar, D. (2018). Validating teacher effects on students' attitudes and behaviors: Evidence from random assignment of teachers to students. *Education Finance and Policy, 13*(3) 281-309.

Blazar, D., & Kraft, M. A. (2017). Teacher and teaching effects on students' attitudes and behaviors. *Educational Evaluation and Policy Analysis, 39*(1), 146-170.

Bolt, D., Wang, Y. C., Meyer, R. H., & Rice, A. B. (2018). IRT mixture model for rating scale confusion associated with negatively worded items. *Paper presented at the National Council on Measurement in Education (NCME) Annual Conference. New York, NY.*

Bound, J., Brown, C., & Mathiowetz, N. (2001). Measurement error in survey data. In J. H. Heckman & E. Learner (Eds.), *Handbook of econometrics* (Vol. 5, pp. 3705-3843). Amsterdam: Elsevier Science.

California Department of Education. (2016). *Preparing all students for college, career, life, and leadership in the 21st century*. Sacramento, CA.

Chamberlain, G. E. (2013). Predictive effects of teachers and schools on test scores, college attendance, and earnings. *Proceedings of the National Academy of Sciences of the United States of America, 110*(43), 17176–82.

Chetty, R., Friedman, J. N., & Rockoff, J. E. (2014). Measuring the impacts of teachers I: Evaluating bias in teacher value-added estimates. *American Economic Review 104*(9), 2593–2632.

DePaoli, J. L., Atwell, M. N., & Bridgeland, J. (2017). *Ready to lead: A national principal survey on how social and emotional learning can prepare children and transform schools.* Washington, DC: Civic Enterprises with Hart Research Associates.

Durlak, J. A., Dymnicki, A. B., Taylor, R., Weissberg, R. P., & Schellinger, K. B. (2011). The impact of enhancing students' social and emotional learning: A meta-analysis of school-based universal interventions. *Child Development 82*(1), 405-432.

Fricke, H., Loeb, S., Meyer, R. H., Rice, A. B., Pier, L., & Hough, H. (2019). *Measuring school contributions to growth in social-emotional learning*. Stanford, CA: Policy Analysis for California Education.

Fuller, W. (1987). *Measurement error models.* Hoboken, N.J.: John Wiley & Sons.

Gehlbach, H., & Hough, H. (2018). *Measuring social emotional learning through student surveys in the CORE districts: A pragmatic approach to validity and reliability.* Stanford, CA: Policy Analysis for California Education.

Gershenson, S. (2016). Linking teacher quality, student attendance, and student achievement. *Education Finance and Policy*, *11*(2), 125-149.

Goldhaber, D., Liddle, S., & Theobald, R. (2013). The gateway to the profession: Assessing teacher preparation programs based on student achievement. *Economics of Education Review, 34*, 29-44.

Hamre, B. K., & Pianta, R. C. (2001). Early teacher-child relationships and the trajectory of children's school outcome through eighth grade. *Child Development, 72*(2), 625-638.

Hanushek, E. A., & Rivkin, S. G. (2010). Generalizations about using value-added measures of teacher quality. *American Economic Review, 100*(2), 267-271.

Jackson, C. K. (2014). *Non-cognitive ability, test scores, and teacher quality: Evidence from 9th grade teachers in North Carolina* (NBER Working Paper No. 18624). Cambridge, MA: National Bureau of Economic Research.

Jackson, K. (2018). What do test scores miss? The importance of teacher effects on non-test-score outcomes. *Journal of Political Economy 126*(5), 2072-2107.

Jennings, J. L., & DiPrete, T. A. (2010). Teacher effects on social and behavioral skills in early elementary school. *Sociology of Education*, *83*(2), 135–159.

Jones, S., Bailey, R., Brush, K., Nelson, B., & Barnes, S. (2016). *What is the same and what is different? Making sense of the "non-cognitive domain:" Helping educators translate research into practice.* Cambridge, MA: Harvard Graduate school of Education. Retrieved from: https://easel.gse.harvard.edu/files/gse-easel-lab/files/taxonomy_handout_0.pdf

Kane, T. J., & Staiger, D. O. (2008). *Estimating teacher impacts on student achievement: An experimental evaluation (NBER Working Paper No. 14607).* Cambridge, MA: National Bureau of Economic Research.

Kane, T.J., McCaffrey, D. F., Miller, T., & Staiger, D.O. (2013). *Have we identified effective teachers? Validating measures of effective teaching using random assignment.* Seattle, WA: Bill and Melinda Gates Foundation MET Project.

Koedel, C. (2008). Teacher quality and dropout outcomes in a large, urban school district. *Journal of Urban Economics, 64,* 560-572.

Koedel, C., Mihaly, K., & Rockoff, J. E. (2015). Value-added modeling: A review. *Economics of Education Review, 47*, 180-195.

Kraft, M. A. (2017). Teacher effects on complex cognitive skills and social-emotional competencies. *Journal of Human Resources*.

Ladd, H. F., & Sorensen, L. C. (2017). Returns to teacher experience: Student achievement and motivation in middle school. *Education Finance and Policy*, *12*(2), 241-279.

Liu, J. & Loeb, S. (2018). *Engaging teachers: Measuring the impact of teachers on student attendance in secondary school* (CEPA Working Paper No. 17-01). Retrieved from http://cepa.stanford.edu/wp17-01

Loeb, S., Christian, M. S., Hough, H., Meyer, R. H., Rice, A. B., & West, M. R. (2019). School differences in social-emotional learning gains: Findings from the first large-scale panel survey of students. *Journal of Educational and Behavioral Statistics, 44*(5), 507-542.

Mashburn, A. J., Pianta, R. C., Hamre, B. K., Downer, J. T., Barbarin, O. A., Bryant, D.,…& Howes, C. (2008). Measures of classroom quality in prekindergarten and children's development of academic, language, and social skills. *Child Development, 79*(3), 732-749.

Masters, G.N. (1982). A Rasch model for partial credit scoring. *Psychometrika*, *47*, 149-174.

McCormick, M. P., Cappella, E., O'Connor, E. E., & McClowry, S. G. (2015). Context matters for social-emotional learning: Examining variation in program impact by dimensions of school climate. *American Journal of Community Psychology, 56*, 101-119.

McKown, C., & Taylor, J. (2018). Introduction to the special issue on social-emotional assessment to guide educational practice. *Journal of Applied Developmental Psychology, 55,* 1-3.

Meyer, R., Wang, Y., & Rice, A. (2018). *Measuring students' social-emotional learning among California's CORE districts: An IRT modeling approach* (PACE Working Paper). Retrieved from http://edpolicyinca.org/sites/default/files/Measuring_SEL_May-2018.pdf

Muraki, E. (1992). *A Generalized Partial Credit Model: Application of an EM algorithm*. (ETS Research Report 92-06). Princeton, NJ: Educational Testing Service.

Nagaoka, J., Farrington, C. A., Ehrlich, S. B., & Heath, R. D. (2015). *Foundations for young adult success: A developmental framework*. Chicago, IL: The University of Chicago Consortium on Chicago School Research.

Ruzek, E. A., Domina, T., Conley, A. M., Duncan, G. J., & Karabenick, S. A. (2015). Using value-added models to measure teacher effects on students' motivation and achievement. *The Journal of Early Adolescence*, *35*(5-6), 852–882.

Schweig, J., Baker, G., Hamilton, L. S., & Stecher, B. M. (2018). Building a repository of assessments of interpersonal, intrapersonal, and higher-order cognitive competencies (Research Report No. 2508). Santa Monica, CA: RAND Corporation. Retrieved from https://www.rand.org/pubs/research_reports/RR2508.html

Transforming Education (2016). *Measuring MESH: Student and teacher surveys curated for the CORE districts.* Retrieved from https://www.transformingeducation.org/wp-content/uploads/2017/04/160406_MeasuringMESH_ForRelease2.pdf

Ujifusa, A. (2018). ESSA promised new gauges of school quality. Does it deliver? *Education* Week, 37(25), 13-14. Retrieved from https://www.edweek.org/ew/articles/2018/04/04/essa-promised-new-gauges-of-school-quality.html

West, M. R., Buckley, K., Krachman, S. B., & Bookman, N. (2018). Development and implementation of student social-emotional surveys in the CORE districts. *Journal of Applied Developmental Psychology* 55, 119-129.

West, M. R., Kraft, M. A., Finn, A. S., Martin, R. E., Duckworth, A. L., Gabrieli, C. F. O., & Gabrieli, J. D. E. (2016). Promise and paradox: Measuring students' non-cognitive skills and the impact of schooling. *Educational Evaluation and Policy Analysis, 38*(1), 148-170.